

Extensión de la SVM para regresión a tramos

David Ramírez, Ignacio Santamaría, Steven Van Vaerenbergh y Javier Vía

{ramirezgd,nacho,steven,jvia}@gtas.dicom.unican.es

Dpto. de Ingeniería de Comunicaciones . Universidad de Cantabria.

39005 Santander, España.

Abstract—In this paper, a new technique for unidimensional regression based on support vector machines is presented. The method is based on the addition of new linear restrictions to the standard (or global) support vector machines technique. Despite the new linear restrictions, a quadratic programming problem is also obtained. Moreover, the use of support vector machines allows us a straightforward derivation of a nonlinear extension of the proposed technique. Finally, simulation results are presented in which the proposed technique is applied to an equalization problem, in this example the proposed technique presents similar results to those of the Bayesian (optimal) equalizer. The technique is also applied to a nonlinear regression problem in which local modeling provides better results than global modeling.

I. INTRODUCCIÓN

Las máquinas de vectores soporte (*Support Vector Machines* o SVM) [1] son un método general para la resolución de problemas de regresión y clasificación. Para regresión, la ventaja de las SVM respecto a otras técnicas es su capacidad para generalizar adecuadamente a partir de un conjunto reducido de patrones de entrenamiento. Esto es así debido a que en su formulación emplea el principio de minimización del riesgo estructural (*Structural Risk Minimization* o SRM) en lugar de la minimización del riesgo empírico (*Empirical Risk Minimization* o ERM) típico, por ejemplo, de las redes neuronales convencionales.

Por otro lado, en algunos problemas de modelado es mejor emplear modelado local en lugar de modelado global. Por ejemplo, para problemas de regresión lineal, un modelo lineal a tramos (*Piecewise-linear* o PWL) presenta muy buenos resultados. En esta línea esta el modelo lineal a tramos canónico de Chua [2] o los hiperplanos bisagra (*Hinging hyperplanes*) de Breiman [3]. Siguiendo esta misma idea de emplear modelado local en lugar de modelado global, pretendemos aplicar el principio SRM en cada tramo forzando la continuidad de nuestro modelo. De esta manera estamos creando una SVM a tramos, lineal o no lineal.

II. SVM CON UN PUNTO DE RUPTURA

En esta sección vamos a presentar un método para regresión a tramos con un único punto de ruptura, x_{br} que de momento asumimos conocido, para un espacio de entrada unidimensional. Para ello, se parte de los patrones de entrenamiento que estarán etiquetados dependiendo del tramo al que pertenezcan:

$$\begin{aligned} (x_i^{(1)}, y_i^{(1)}), & \quad \text{si } x_i \leq x_{br}, & \forall i = 1, \dots, N_1, \\ (x_i^{(2)}, y_i^{(2)}), & \quad \text{si } x_i > x_{br}, & \forall i = 1, \dots, N_2. \end{aligned}$$

En primer lugar se presentará el método para el caso de regresión lineal y a continuación se hará la extensión a regresión no lineal.

A. Regresión Lineal: PWL-SVM

Para el caso lineal nuestro modelo viene dado por

$$y = f(x) = \begin{cases} w_1 x + b_1, & \text{si } x \leq x_{br} \\ w_2 x + b_2, & \text{si } x > x_{br} \end{cases},$$

además se debe cumplir que el modelo sea continuo, es decir, se debe cumplir

$$w_1 x_{br} + b_1 = w_2 x_{br} + b_2.$$

Siguiendo el principio SRM para el ajuste del modelo debemos minimizar la siguiente función de coste:

$$J = \frac{1}{2} \sum_{k=1}^2 \|w_k\|^2 + C \sum_{k=1}^2 \sum_{i=1}^{N_k} (\xi_i^{(k)} + \tilde{\xi}_i^{(k)}), \quad (1)$$

sujeto a las siguientes restricciones

$$w_k x_i^{(k)} + b_k - y_i^{(k)} \leq \epsilon + \xi_i^{(k)}, \quad \forall i = 1, \dots, N_k \quad (2)$$

$$y_i^{(k)} - w_k x_i^{(k)} - b_k \leq \epsilon + \tilde{\xi}_i^{(k)}, \quad \forall i = 1, \dots, N_k \quad (3)$$

$$w_1 x_{br} + b_1 = w_2 x_{br} + b_2, \quad (4)$$

$$\xi_i^{(k)}, \tilde{\xi}_i^{(k)} \geq 0, \quad (5)$$

con $k = 1, 2$.

Se ha asumido como función de perdidas la función ϵ -insensitive de Vapnik. La principal diferencia respecto a la SVM tradicional (o global) es que se ha introducido una restricción lineal más dada por (4) que, como ya se ha mencionado antes, fuerza la continuidad del modelo. Para resolver el problema de optimización anterior recurrimos a los multiplicadores de Lagrange, en primer lugar creamos el Lagrangiano

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \sum_{k=1}^2 \|w_k\|^2 + C \sum_{k=1}^2 \sum_{i=1}^{N_k} (\xi_i^{(k)} + \tilde{\xi}_i^{(k)}) \\ & - \sum_{k=1}^2 \sum_{i=1}^{N_k} \alpha_i^{(k)} (\epsilon + \xi_i^{(k)} + y_i^{(k)} - w_k x_i^{(k)} - b_k) \\ & - \sum_{k=1}^2 \sum_{i=1}^{N_k} \tilde{\alpha}_i^{(k)} (\epsilon + \tilde{\xi}_i^{(k)} - y_i^{(k)} + w_k x_i^{(k)} + b_k) \\ & + \lambda \{(w_1 - w_2) x_{br} + (b_1 - b_2)\}, \quad (6) \end{aligned}$$

donde $\alpha_i^{(k)}$, $\tilde{\alpha}_i^{(k)}$, $r_i^{(k)}$ y $\tilde{r}_i^{(k)}$ son los multiplicadores de Lagrange, que deben ser positivos. La solución al problema (6) viene dada por el punto de ensilladura del Lagrangiano, es decir, que debemos minimizar (6) respecto a w_k , b_k , $\xi_i^{(k)}$ y $\tilde{\xi}_i^{(k)}$ con $k = 1, 2$ y maximizarlo respecto a los multiplicadores de Lagrange.

Tomando derivadas parciales, igualándolas a 0 y sustituyendo en (6) obtenemos el problema dual de optimización (7). Dicho problema, al igual que en la obtención de la SVM global, es un problema de programación cuadrática (*quadratic programming* o QP).

$$\begin{aligned} \argmin_{\alpha_i^{(k)}, \tilde{\alpha}_i^{(k)}} & \frac{1}{2} \sum_{k=1}^2 \sum_{i,j=1}^{N_k} [(\tilde{\alpha}_i^{(k)} - \alpha_i^{(k)}) (\tilde{\alpha}_j^{(k)} - \alpha_j^{(k)}) \times \\ & (x_i^{(k)} - x_{br}) (x_j^{(k)} - x_{br})] \\ & + \epsilon \sum_{k=1}^2 \sum_{i=1}^{N_k} (\tilde{\alpha}_i^{(k)} + \alpha_i^{(k)}) - \sum_{k=1}^2 \sum_{i=1}^{N_k} y_i^{(k)} (\tilde{\alpha}_i^{(k)} - \alpha_i^{(k)}), \quad (7) \end{aligned}$$

sujeto a

$$\begin{aligned} \sum_{k=1}^2 \sum_{i=1}^{N_k} (\tilde{\alpha}_i^{(k)} - \alpha_i^{(k)}) &= 0, \\ \alpha_i^{(k)}, \tilde{\alpha}_i^{(k)} &\in [0, C], \quad \forall i = 1, \dots, N_k. \end{aligned}$$

Si expresamos (7) de manera matricial, nuestro problema es

$$\tilde{\alpha}, \alpha = \underset{\tilde{\alpha}, \alpha}{\operatorname{argmin}} \frac{1}{2} (\tilde{\alpha} - \alpha)^T \mathbf{K} (\tilde{\alpha} - \alpha) + \epsilon (\tilde{\alpha} + \alpha)^T \mathbf{1} - (\tilde{\alpha} - \alpha)^T \mathbf{y}, \quad (8)$$

sujeto a las mismas restricciones, donde $\mathbf{y} = (y_1^{(1)}, \dots, y_{N_1}^{(1)}, y_1^{(2)}, \dots, y_{N_2}^{(2)})^T$, α y $\tilde{\alpha}$ se definen análogamente a \mathbf{y} , $\mathbf{1}$ es un vector de las dimensiones adecuadas con todas sus componentes iguales a uno, y por último la matriz de *kernel*, que es diagonal por bloques, se define de la siguiente manera

$$\mathbf{K} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix}$$

siendo

$$[\mathbf{A}_k]_{i,j} = (x_i^{(k)} - x_{br}) (x_j^{(k)} - x_{br}), \quad \forall i, j = 1, \dots, N_k$$

Una vez resuelto el problema QP, podemos obtener los w_k de la siguiente manera

$$w_k = \sum_{i=1}^{N_k} (\tilde{\alpha}_i^{(k)} - \alpha_i^{(k)}) (x_i^{(k)} - x_{br}).$$

Finalmente el cálculo de los b_k se puede hacer mediante las condiciones de Karush-Kuhn-Tucker (KKT) [1]:

$$\begin{aligned} b_k &= y_i^{(k)} - w_k x_i^{(k)} + \epsilon, \quad \text{si } \alpha_i^{(k)} \in (0, C) \text{ y } \tilde{\alpha}_i^{(k)} = 0 \\ b_k &= y_i^{(k)} - w_k x_i^{(k)} - \epsilon, \quad \text{si } \tilde{\alpha}_i^{(k)} \in (0, C) \text{ y } \alpha_i^{(k)} = 0 \end{aligned}$$

con $1 \leq i \leq N_k$.

Es importante destacar que una vez resuelto el problema y obtenidos los w_k y los b_k no es necesario conocer los $x_i^{(k)}$ para calcular la salida del modelo. Esto no ocurre en el caso de regresión no lineal.

B. Regresión no Lineal: PWNL-SVM

La extensión a un modelo no lineal a tramos (*Piecewise Non Linear* o PWNL) es prácticamente directa aplicando el denominado *kernel trick*. En primer lugar los datos de entrada son transformados no linealmente a un espacio de mayor dimensión, inclusive infinito, mediante la transformación $\phi(\cdot)$,

$$\mathbb{R} \xrightarrow{\phi(\cdot)} \mathbb{R}^H,$$

y en dicho espacio transformado aplicamos el algoritmo lineal. Por lo tanto, en el espacio transformado el modelo viene dado por

$$y = f(x) = \begin{cases} \mathbf{w}_1^T \phi(x) + b_1, & \text{si } x \leq x_{br} \\ \mathbf{w}_2^T \phi(x) + b_2, & \text{si } x > x_{br} \end{cases},$$

sujeto a la siguiente restricción

$$\mathbf{w}_1^T \phi(x) + b_1 = \mathbf{w}_2^T \phi(x) + b_2.$$

Aplicando los mismos pasos que en el caso lineal, obtenemos un problema de optimización equivalente a (8), con la diferencia que en este caso los bloques de la matriz de *kernel* vienen dados por

$$[\mathbf{A}_k]_{i,j} = (\phi(x_i^{(k)}) - \phi(x_{br}))^T (\phi(x_j^{(k)}) - \phi(x_{br}))$$

con $1 \leq i, j \leq N_k$.

Definido de esta manera debemos conocer la transformación no lineal $\phi(\cdot)$. Sin embargo, mediante el *kernel trick*, es posible sustituir los productos escalares en el espacio transformado por una función *kernel*. Dicha función *kernel* debe satisfacer unas condiciones conocidas como las condiciones de Mercer [1]. Finalmente, los bloques de la matriz de *kernel* se pueden expresar de la siguiente manera

$$[\mathbf{A}_k]_{i,j} = \kappa(x_i^{(k)}, x_j^{(k)}) - \kappa(x_{br}, x_j^{(k)}) - \kappa(x_i^{(k)}, x_{br}) + \kappa(x_{br}, x_{br}), \quad \forall i, j = 1, \dots, N_k.$$

Por otro lado, el cálculo de los b_k también se hace a partir de las condiciones KKT:

$$\begin{aligned} b_k &= y_j^{(k)} - \sum_{j=1}^{N_k} (\tilde{\alpha}_j^{(k)} - \alpha_j^{(k)}) [\kappa(x_j^{(k)}, x_i^{(k)}) \\ &\quad - \kappa(x_{br}, x_i^{(k)})] + \epsilon, \quad \text{si } \alpha_j^{(k)} \in (0, C) \text{ y } \tilde{\alpha}_j^{(k)} = 0, \\ b_k &= y_j^{(k)} - \sum_{j=1}^{N_k} (\tilde{\alpha}_j^{(k)} - \alpha_j^{(k)}) [\kappa(x_j^{(k)}, x_i^{(k)}) \\ &\quad - \kappa(x_{br}, x_i^{(k)})] - \epsilon, \quad \text{si } \tilde{\alpha}_j^{(k)} \in (0, C) \text{ y } \alpha_j^{(k)} = 0, \end{aligned}$$

con $1 \leq i \leq N_k$. Una vez resuelto el problema QP y obtenidos los b_k , la salida de nuestro modelo vendrá dada por

$$y = \sum_{i=1}^{N_k} (\tilde{\alpha}_i^{(k)} - \alpha_i^{(k)}) [\kappa(x_i^{(k)}, x) - \kappa(x_{br}, x)] + b_k,$$

donde $k = 1$ si $x \leq x_{br}$ y $k = 2$ si $x > x_{br}$.

Finalmente cabe destacar que es posible aplicar transformaciones diferentes a cada tramo, obteniendo dos funciones *kernel* diferentes. Esta característica facilita la aplicación de modelado local en lugar de modelado global.

III. GENERALIZACIÓN A M PUNTOS DE RUPTURA

En esta sección se generaliza la técnica de regresión unidimensional al caso en el que tenemos M puntos de ruptura, que además deben ser conocidos: $x_{br_1}, \dots, x_{br_M}$. Se presentará la técnica únicamente para el caso lineal, para realizar su extensión a regresión no lineal basta con sustituir los productos internos en el espacio de entrada por productos internos en el espacio transformado y aplicar el *kernel trick*.

Siguiendo un desarrollo similar al que se ha realizado en la sección anterior, pero añadiendo una restricción adicional por cada punto de ruptura obtenemos un problema análogo a (8) en el que la matriz de *kernel*, de dimensiones $\sum_{k=1}^{M+1} N_k \times \sum_{k=1}^{M+1} N_k$, se calcula de la siguiente manera¹

$$\mathbf{K} = \mathbf{A} + \sum_{k=1}^{M-1} (\mathbf{B}_k + \mathbf{B}_k^T) + \sum_{k=1}^{M-1} \mathbf{C}_k, \quad (9)$$

siendo \mathbf{A} una matriz diagonal por bloques con la siguiente estructura

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \dots & \mathbf{0} \\ \vdots & \dots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_{M+1} \end{pmatrix},$$

y las \mathbf{A}_k vienen dadas por

$$[\mathbf{A}_k]_{i,j} = (x_i^{(k)} - x_{br_{k-1}})^T (x_j^{(k)} - x_{br_{k-1}})$$

con $1 \leq i, j \leq N_k$. Cabe mencionar que por simplificar la notación, se ha añadido un punto de ruptura adicional ficticio $x_{br_0} = x_{br_1}$. Por otro lado, la definición de las matrices \mathbf{B}_k y \mathbf{C}_k viene dada por (10) y (11), respectivamente. Para finalizar, $\mathbf{y} = (y_1^{(1)}, \dots, y_{N_1}^{(1)}, y_1^{(2)}, \dots, y_{N_{M+1}}^{(M+1)})^T$, α y $\tilde{\alpha}$ se definen

¹Es importante resaltar que si el índice inferior de un sumatorio es mayor que el superior, dicho sumatorio vale 0.

$$[\mathbf{B}_k]_{i,j} = \begin{cases} (x_{br_{k+1}} - x_{br_k}) (x_j^{(k+1)} - x_{br_k}), & \text{si } \sum_{l=1}^{k+1} N_l < i \leq \sum_{l=1}^{M+1} N_l \text{ y } \sum_{l=1}^k N_l < j \leq \sum_{l=1}^{k+1} N_l \\ 0, & \text{en caso contrario} \end{cases} \quad (10)$$

$$[\mathbf{C}_k]_{i,j} = \begin{cases} (x_{br_k} - x_{br_{k+1}}) (x_{br_k} - x_{br_{k+1}}), & \text{si } \sum_{l=1}^{k+1} N_l < i, j \leq \sum_{l=1}^{M+1} N_l \\ 0, & \text{en caso contrario} \end{cases} \quad (11)$$

análogamente a \mathbf{y} , y los w_k se calculan de la siguiente manera

$$w_1 = \sum_{i=1}^{N_1} (\tilde{\alpha}_i^{(1)} - \alpha_i^{(1)}) (x_i^{(1)} - x_{br_1})$$

$$w_k = \sum_{i=1}^{N_k} (\tilde{\alpha}_i^{(k)} - \alpha_i^{(k)}) (x_i^{(k)} - x_{br_{k-1}}) +$$

$$(x_{br_k} - x_{br_{k-1}}) \sum_{l=k+1}^{M+1} \sum_{i=1}^{N_l} (\tilde{\alpha}_i^{(l)} - \alpha_i^{(l)}), \forall k = 2, \dots, M+1$$

IV. RESULTADOS DE SIMULACIÓN

En esta sección vamos a presentar diversos resultados de simulación de la técnica de regresión que se ha propuesto. En primer lugar lo aplicaremos a un problema de igualación no ciega en el que un modelo PWL-SVM trata de aproximar la frontera óptima de decisión. En segundo lugar compararemos la SVM no lineal a tramos (PWNL-SVM) con una SVM global.

A. Igualación no ciega

En este apartado vamos a aplicar una SVM lineal a tramos al problema de igualación no ciega de sistemas SISO (*Single-Input Single-Output*). Comenzaremos por realizar un breve introducción a la igualación no ciega y después veremos como aplicar la técnica propuesta a dicho problema.

Si asumimos un sistema de comunicaciones digitales muestreado a la tasa de símbolo, la señal a la entrada del igualador es

$$y[n] = \sum_{k=0}^N h[n-k]s[k] + e[n], \quad (12)$$

donde $h[n]$ es el canal equivalente, que se ha asumido FIR con $N+1$ coeficientes. Por otro lado, $s[n]$ son los símbolos transmitidos que asumiremos que pertenecen a una modulación BPSK ($s[n] \in \{1, -1\}$) y $e[n]$ es un ruido aditivo blanco y Gaussiano (AWGN) con varianza σ_e^2 .

Bajo estas condiciones el igualador óptimo símbolo a símbolo es el igualador Bayesiano. Para un igualador transversal de longitud M , la estima del símbolo transmitido se realiza de la siguiente manera

$$\hat{s}[n-d] = \text{signo}(g(\mathbf{y}[n]))$$

donde $\mathbf{y}[n] = (y[n], \dots, y[n-M+1])^T$ y $g(\cdot)$ es una función discriminante, lineal o no lineal, de las observaciones. En [4] se propuso que el problema de igualación óptima puede verse como un problema de clasificación entre los estados del canal. Vamos a explicar un poco más en detalle esta idea, para lo cual empezaremos por expresar (12) de manera matricial en una situación sin ruido

$$\mathbf{r}[n] = \mathbf{H}\mathbf{s}[n]$$

donde $\mathbf{s}[n] = (s[n], \dots, s[n-M-N+1])^T$ y \mathbf{H} es la matriz de convolución del canal, de dimensiones $M \times (M+N)$ y estructura Toeplitz, que viene dada por

$$\mathbf{H} = \begin{pmatrix} h[0] & \dots & h[N] & 0 & \dots & 0 \\ 0 & h[0] & \dots & h[N] & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h[0] & \dots & h[N] \end{pmatrix}.$$

Dadas las dimensiones del vector $\mathbf{s}[n]$ y el tipo de modulación (BPSK), existen 2^{M+N} vectores \mathbf{s}_i posibles, de manera que también existen 2^{M+N} estados del canal diferentes (\mathbf{r}_i). Dichos estados del canal se pueden obtener de la siguiente manera: $R = \{\mathbf{r}_i = \mathbf{H}\mathbf{s}_i, \forall i = 1, \dots, 2^{M+N}\}$. Para una modulación BPSK, R se puede dividir en dos subconjuntos

$$R^{(\pm 1)} = \{\mathbf{r}_i, s[n-d] = \pm 1\}.$$

Con lo cual, el igualador deberá encontrar la frontera de clasificación entre $R^{(+)}$ y $R^{(-)}$: $g(\mathbf{y}[n]) = 0$. En particular, la frontera óptima, dada por el igualador Bayesiano, es no lineal y viene dada por

$$g(\mathbf{y}[n]) = \sum_{\mathbf{r}_i \in R^{(+1)}} \exp\left(-\frac{\|\mathbf{y}[n] - \mathbf{r}_i\|^2}{2\sigma_e^2}\right) - \sum_{\mathbf{r}_i \in R^{(-1)}} \exp\left(-\frac{\|\mathbf{y}[n] - \mathbf{r}_i\|^2}{2\sigma_e^2}\right) \quad (13)$$

Por otro lado, la frontera que se obtiene mediante un igualador lineal viene dada por

$$g(\mathbf{y}[n]) = \mathbf{w}^T \mathbf{y}[n]$$

donde \mathbf{w} se suele obtener de acuerdo a un criterio de cero forzado (*zero forcing* o ZF), o de acuerdo a un criterio de mínimo error cuadrático medio (*minimum mean square error* o MMSE). Si se emplea el criterio MMSE se obtiene el filtro de Wiener [5].

Como se puede ver la función discriminante óptima entre los estados del canal (13), es relativamente compleja y el gasto computacional asociado puede ser elevado. Para disminuir dicha complejidad computacional han aparecido diversas técnicas, algunas de ellas basadas en SVM [6], [7]. Lo que se propone en este artículo es aproximar la frontera óptima de decisión ($g(\mathbf{y}[n]) = 0$) mediante tramos lineales, de manera que el gasto computacional sea menor una vez entrenada la PWL-SVM. Dado que el método desarrollado es para regresión unidimensional estamos limitados al caso con $M = 2$. Para entrenar la PWL-SVM emplearemos L puntos de la frontera y el decisor viene dado por $\hat{s}[n-d] = \text{signo}(y[n-1] - f(y[n]))$, donde $f(y[n])$ es nuestra función PWL-SVM. Por el momento, el número de puntos de ruptura de la aproximación PWL de la frontera óptima hay que elegirlo *a priori*, aunque si se sobreestima las prestaciones son similares. Por otro lado se ha asumido conocido el valor de dichos puntos de ruptura en los desarrollos efectuados en las secciones II y III, sin embargo, para esta aplicación es necesario calcularlos, para ello emplearemos el método propuesto por Chua [2]. Dicho método se basa en minimizar iterativamente el error cuadrático medio mediante búsqueda lineal en la dirección dada por un método de segundo orden.

Para este ejemplo ilustrativo del método propuesto hemos empleado un canal extremadamente sencillo dado por $H(z) = 1 + 0.5z^{-1}$, el número de puntos de ruptura elegido es 2 y el retraso $d = 1$. En primer lugar, en la figura 1 se puede ver la frontera óptima de decisión, la frontera de PWL-SVM (para su entrenamiento se han usado $L = 50$ puntos de la frontera óptima), la frontera del igualador MMSE y los estados del canal. Como se puede ver, la aproximación de la frontera óptima por la PWL-SVM es muy buena. Finalmente, para comparar las tres técnicas de igualación hemos obtenido la BER,

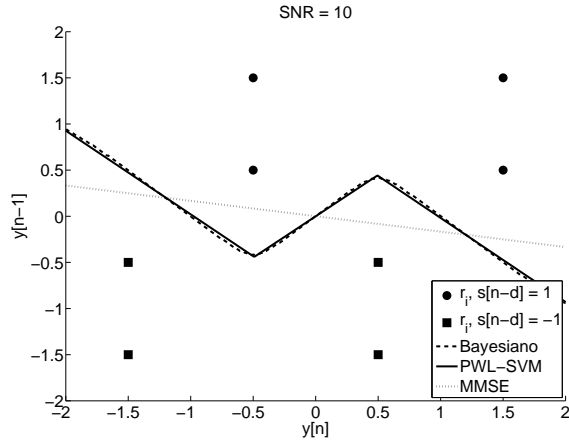


Fig. 1. Fronteras de decisión.

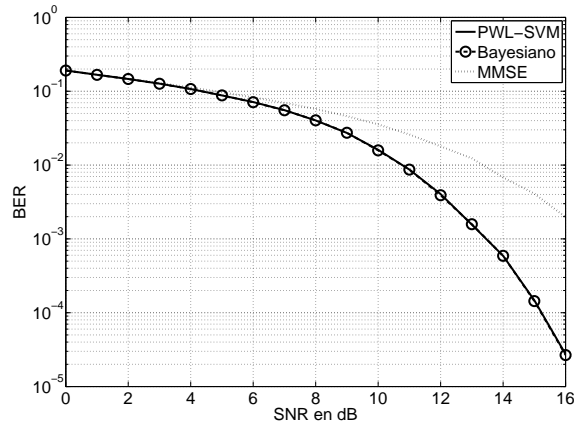


Fig. 2. BER para las distintas técnicas de igualación.

que se puede ver en la figura 2. En dicha figura se puede observar como la igualación mediante PWNL-SVM presenta unos resultados muy próximos al igualador Bayesiano y como el MMSE se queda lejos de ambos. Es decir, que obtenemos unos resultados próximos al igualador óptimo con un incremento computacional despreciable respecto al igualador MMSE, sólo es necesario realizar una serie de comparaciones para ver en qué tramo se encuentra $y[n]$.

B. Modelado global vs. local

En este apartado comparamos la SVM no lineal a tramos (*piecewise-nonlinear-SVM* o PWNL-SVM) con una SVM no lineal global. Lo que se quiere demostrar es que en algunos casos, el aplicar modelado local en lugar de global presenta mejores resultados en prestaciones y en la complejidad del modelo final resultante. Para ello hemos generado un modelo que presenta dos zonas diferentes ($x \geq 4$), una de variaciones bruscas y otra de variaciones más suaves. Para entrenar las SVM se emplean muestras aleatoriamente seleccionadas del modelo. En la figura 3 se puede ver el modelo teórico, las muestras empleadas para el entrenamiento (40 en total), el resultado de la PWNL-SVM y la SVM no lineal global, ambas con *kernel* Gaussiano. Los parámetros elegidos son: $C = 20$, $\epsilon = 0.001$, $\sigma^2 = 0.5$ para la PWNL en la zona de cambios abruptos y $\sigma^2 = 1$ en la zona de cambios suaves, para la SVM global se han usado también los mismos σ^2 , uno para cada una de las SVM entrenadas. En dicha figura se observa como la PWNL-SVM presenta unos resultados superiores a la SVM no lineal global usando entre seis u ocho vectores soporte menos: 31 de la PWNL-SVM frente a 37 y 39

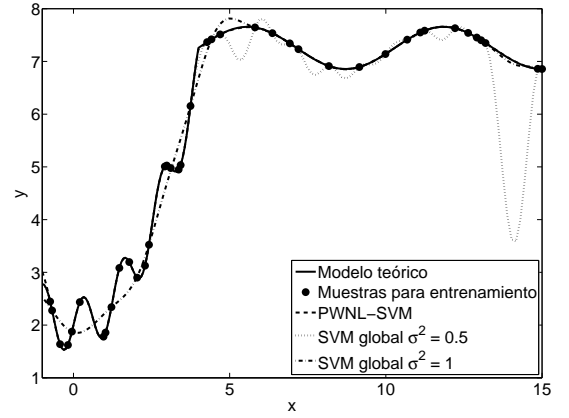


Fig. 3. Modelado global vs. local.

de la SVM no lineal global para cada σ^2 , respectivamente. El modelo PWNL-SVM prácticamente es igual al modelo teórico. Para explicar la mejora obtenida es fácil ver que para la SVM global si usamos una varianza del *kernel* elevada no puede ajustar adecuadamente la zona de variaciones bruscas y si su varianza es pequeña ajusta mal la zona de variaciones suaves. Esto no ocurre con la PWNL-SVM ya que nos permite usar dos varianzas diferentes, una por zona.

V. CONCLUSIONES

En este artículo se ha presentado una técnica de regresión a tramos mediante SVM. Esto se ha conseguido añadiendo una restricción adicional al problema por cada punto de ruptura, resultando al final un problema QP al igual que en la SVM tradicional. Sin embargo, mediante la técnica propuesta la extensión a regresión multidimensional no es inmediata. Por otro lado, al usar las SVM la extensión a regresión no lineal es inmediata sin más que aplicar el *kernel trick*, y en general son capaces de generalizar mejor que otras técnicas. Por otro lado se han presentado dos aplicaciones de la técnica de regresión propuesta. En primer lugar se ha aplicado a igualación no ciega aplicando regresión sobre la frontera óptima de decisión, presentando resultados muy cercanos al igualador Bayesiano. Y en segundo lugar se ha aplicado a un problema de regresión no lineal, en el cual presenta ciertas ventajas respecto a una regresión global.

AGRADECIMIENTOS

Este trabajo ha sido financiado por el MEC (Ministerio de Educación y Ciencia) bajo el contrato TEC2004-06451-C05-02 y el programa de becas de formación de profesorado universitario (AP2005-5366 y AP2006-2965).

REFERENCES

- [1] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.
- [2] L. O. Chua and A. Deng, "Canonical piecewise-linear modeling," *IEEE Trans. on Circuits and Systems*, vol. 33, no. 5, pp. 511–525, May 1986.
- [3] L. Breiman, "Hinging hyperplanes for regression, classification and function approximation," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 999–1013, May 1993.
- [4] G. J. Gibson, S. Siu, and C. F. Cowan, "The application of nonlinear structures to the reconstruction of binary signals," *IEEE Transactions on Signal Processing*, vol. 39, no. 8, pp. 1877–1884, August 1991.
- [5] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2002.
- [6] F. Pérez-Cruz, P. Alarcón-Diana, A. Navia-Vázquez, and A. Artés-Rodríguez, "SVC-based equalizer for burst TDMA transmissions," *Signal Processing*, vol. 81, no. 8, pp. 1681 – 1693, August 2001.
- [7] I. Santamaría, R. González, C. Pantaleón, and J. C. Príncipe, "Maximum margin equalizers trained with the Adatron algorithm," *Signal Processing*, vol. 83, no. 3, pp. 593–602, 2003.